



Position Paper for the Use of PDF/A in Management of Quality Records

PP05-2009

NUCLEAR INFORMATION AND RECORDS MANAGEMENT ASSOCIATION
10 Almas Road, Windham, NH 03087

Nuclear Information and Records Management Association (NIRMA) represent a consensus of the members of the committee, subcommittee, or ad hoc committee with the concurrence of the NIRMA membership and Board of Directors. Guidelines are recommended procedures for establishment of a program or performance of a task related to the interests of NIRMA members. Guidelines represent the consensus of the sponsoring committee on the best information and method of the subject based upon applicable regulatory requirements, industry standards and practical experience. A lesser amount of detail may be used to identify the areas of interface outside the scope of the document. Guidelines are not intended to be interpreted as industry standards but as suggested guidance with the understanding that companies are not obligated to use the suggested guidance.

NOTICE

This document was prepared by a committee of the Nuclear Information and Records Management Association (NIRMA). Neither this committee, NIRMA, members of NIRMA, other persons contributing to or assisting in the preparation of the document, nor any person acting on behalf of these parties (a) makes any warranty or representation, expressed or implied, with respect to the accuracy, completeness, or usefulness of the information contained in this document, or that the use of any information, apparatus, method, or process disclosed in this document may not infringe on privately owned rights, or (b) assumes liabilities with respect to the use of any information, apparatus, method, or process disclosed in this document.

©2009. Nuclear Information and Records Management Association. All rights reserved.

CONTENTS

Foreword.....iv

Abstract v

1.0 Introduction..... 1

2.0 ISO 19005-1, PDF/A Standard 2

 2.1 Purpose of PDF/A Usage 3

 2.2 Advantages of the PDF/A Format Type 3

 2.3 What PDF/A Allows and Disallows 3

3.0 Regulatory Requirements 4

4.0 NIRMA’s Position..... 4

 4.1 General Endorsement 4

 4.2 NARA Requirements Met..... 4

 4.3 NRC Requirements Met 5

 4.4 Implementation Guidance 6

5.0 References 6

Foreword

NIRMA provides this paper as a supplement to NIRMA TG15, “Management of Electronic Records” to help organizations evaluate the usefulness of implementing a records system using the sustainable form at of PDF/A.

NIRMA wishes to acknowledge the efforts of the following subcommittee members for their work and significant contributions:

Drafting Committee

Alan Rabe
FirstEnergy Corporation

Lona Smith
South Texas Project

Eugene Yang
KISMET Consulting, Inc.

Abstract

This document endorses the use of the ISO 19005 PDF/A standard as a sustainable file format type for use in the nuclear industry. It provides additional information for consideration in its use for site quality records and suggests standard settings for meeting the NRC submittal requirements.

1.0 Introduction

Most nuclear facilities' retention policies were built on regulatory standards which state that "lifetime quality assurance records" are required to be maintained for the life of the particular item while it is installed in the plant. These were also called "permanent" records. Originally, this meant retaining records for approximately 40 years. However, for example, with license renewal and plant life extension activities, another 20 years of operating life is being added onto these facilities, thus extending the life-of-plant retention period. In addition, for certain record types, other requirements further extend the retention period and may exceed 100 years (e.g., personnel radiation protection records, asbestos exposure, and environmental data).

Many electronic documents are created using formats that are proprietary and not suitable for long-term preservation ("sustainable"). Some sustainable formats are no longer considered optimum; for example, an ASCII text file that does not retain its format structure. NIRMA TG-15, "Management of Electronic Records," states: A record is "Information regardless of physical form or characteristics, appropriate for preservation as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities of the organization. Examples of where this information may reside are: books, papers, maps, photographs, machine readable electronic files, or other documentary materials." Many of these records needed a new format to ensure that they could be retained, retrieved, and adequately stored for their retention period. Now, for most plants, the "file format of choice" for electronic permanent plant records is Adobe Corporation's Portable Document Format (PDF), stored on magnetic hard drive, compact disc, or optical disk media. This product provides a common, generic format while integrating with common office applications; today, most office desktop applications can create PDF output.

PDF is a digital format for representing documents. PDF files may be:

- Created from native applications (e.g., Microsoft Word, Excel, AutoCAD)
- Converted from other electronic formats (e.g., multi-page Tagged Image File Format [TIFF]), or
- Digitized from paper or other hardcopy media (e.g., through scanners and scanning software).

The primary question becomes: Can these electronic records be read 50 to 60, or 100+ years from now...or beyond?

Given that the format of choice is PDF, the primary concern is the consistent generation of electronic QA records that dictate the ability of users in the far future to open and read these records. The implementation of the Adobe Acrobat application at the various nuclear facilities has ranged from a controlled server to individual desktop deployments. In the case of personal computer installations, configuration settings can be (and are) set by the individual user, and therefore, create PDFs based on a wide range of "settings" (e.g., resolution, compression, color settings, linearization, etc.). The ever-

increasing functionality from Adobe raises the specter of greater variance in configuration settings, which further increases the future tension of being able to open the record in the future.

New technological features in the future development of Adobe's products increase the uncertainty that important documents and records being produced today can be read in the future. For example, you can embed video formats in today's PDF files, but will tomorrow's PDF readers be able to open today's video technology when it becomes antiquated? Also, reliance on Adobe implies that an electric utility/power plant is being subjected to a particular software vendor's direction. The concern is there have been numerous examples of software becoming an industry de facto standard, only to see the software become defunct. Hence, a different approach is suggested, that is based on standards, for storing records in a long-term format.

2.0 ISO 19005-1, PDF/A Standard

Because PDF files are being kept permanently, they must remain usable and accessible across multiple generations of technology. In 2005, the International Standards Organization (ISO) formally approved and issued ISO 19005-1, "Document Management – Electronic Document File Format for Long-term Preservation – Part 1: Use of PDF 1.4 (PDF/A-1)," which establishes PDF/A ("PDF-Archive") as an output format that is self-descriptive. This means that a "reader" based on the specifications described in the ISO standard can be developed that will be able to open these files. A key element is the requirement for PDF/A documents to be 100% self-contained. All of the information necessary for displaying the document in the same manner every time is embedded in the file. This includes all visible content like text, raster images, vector graphics, fonts, color information. A PDF/A document is not permitted to be reliant on any information from external sources, for example links to external image files or fonts that are not embedded. The standard provides:

- Data format for static content and accurate image rendering
- Framework for capturing of metadata
- Framework for creation of logical document structure

The standard cites two acceptable levels of compliance for PDF/A files:

- PDF/A-1b includes all the requirements of the standard and provides the conformance level to ensure the rendered visual appearance is preserved over the long term.
- PDF/A-1a includes all the requirements of PDF/A-1b and additionally requires that document structure be included (also known as being "tagged"), with the objective of ensuring that document content can be searched and repurposed.

PDF/A is a flavor of PDF – any documents created using the PDF/A standard will have the file extension of “.pdf” (e.g., mydocument.pdf). The standard essentially restricts job option features in the creation of a PDF so that it can be readable in the future.

2.1 Purpose of PDF/A Usage

The primary purpose of PDF/A, as identified in the standard, is to provide “a mechanism for representing electronic documents in a manner that preserves their visual appearance over time, independent of the tools and systems used for creating, storing, or rendering the files.” The standard is based on the PDF 1.4 specification, eXtensible Metadata Platform (XMP), and ICC (International Color Consortium) standards. Other characteristics of PDF/A include:

- Device Independent - Can be reliably and consistently rendered without regard to the hardware or software platform.
- Self-contained – Contains all resources necessary for rendering, e.g., embedded fonts
- Self-documenting – Contains its own description
- Unfettered – Absence of technical file protection mechanisms
- Available – Authoritative specification publicly available
- Adoption – Widespread use may be the best deterrent against preservation risk

2.2 Advantages of the PDF/A Format Type

Advantages of using the PDF/A standard include:

- Software vendor independence
- PDF/A can include: hidden text, indexing, new compression scheme, and metadata, etc.
- Delivers a unified archival format
- Use of color, in addition to black and white
- Digital Documents – vector graphics
- Data Stream: eXtensible Markup Language (XML), Advanced Function Printing (AFP), IBM Line Printer (1403), Printer Command Language (PCL)
- Accessible through one viewer
- Reduces future migration and management costs
- Supports digital authentication

2.3 What PDF/A Allows and Disallows

PDF/A-1 files must include embedded fonts, device-independent color, and XMP metadata. PDF/A-1 files may NOT include encryption, Lempel-Ziv-Welch algorithm (LZW) compression, embedded files, certain external content references (e.g., certain types of hyperlinks), PDF transparency, multi-media, and JavaScript.

3.0 Regulatory Requirements

Record keeping must meet the requirements of regulatory bodies. The key agency is the US National Archives and Records Administration. Other branches of government adopt standards based on their direction to them.

PDFs need to be created to meet the electronic submission requirements dictated by the U.S. Nuclear Regulatory Commission's (USNRC) "Electronic Maintenance and Submission of Information" Rule (also known as the E-Rule). Appendix A of the E-Rule, "United States Nuclear Regulatory Commission (NRC), Guidance for Electronic Submissions to the Commission," provides specific settings while using Adobe Acrobat (5.0 and later), as well as segment (or "chunk") submissions based on a 50 Megabyte limitation in file size. The E-Rule states that electronic submittals are voluntary. However, because of federal e-Government initiatives, as well as the Government Paper Elimination Act [GPEA], the USNRC has been encouraging licensees to provide submittals electronically, and in some cases, has made it mandatory (e.g., next generation plant early site permit or combined operating license applications).

4.0 NIRMA's Position

4.1 General Endorsement

NIRMA encourages the use of PDF/A to ensure electronic records sustainability.

4.2 NARA Requirements Met

The National Archives and Records Administration (NARA) has determined the following criteria for digital file sustainability (preservation):

- Publicly and openly documented
 - Existence of complete documentation
- Non-proprietary
 - Transparency – degree to which the digital representation is open to direct analysis
 - External Dependencies – degree to which the format is dependent upon specific hardware, operating system, or software for rendering or use and the complexity of dealing with those dependencies in future technical environments
 - Impact of Patents – degree to which the ability of archival institutions to sustain content in a format will be inhibited by patents
 - Technical Protection Mechanisms – implementation of mechanisms that prevent the preservation of content by a trusted authority

- In widespread use
 - degree to which the format is already in use
- Self-documenting
 - digital objects that contain basic descriptive, technical, and other administrative metadata
- Can be opened, read, and accessed with readily available tools

PDF/A meets all of these criteria:

- Publicly and openly documented - by virtue of the process of becoming an international standard, vetted by government agencies, commercial business, educational institutions and other standards organizations
- Non-proprietary – no one company owns the specification for PDF/A; Adobe released the full content of the PDF 1.4 specification to mitigate the proprietary nature of PDF.
- In widespread use – PDF is widely used in all applications, ranging from recordkeeping to publications, in all industries, in all governmental agencies, and educational institutions. As stated earlier, PDF/A is a flavor of PDF and can be widely read by any current PDF reader.
- Self-documenting – one of the keystones of PDF/A is the descriptive nature to allow the ability to “read back” the form and format tagging through the use of a viewer based on the standard.
- Can be opened, read, and accessed with readily available tools – the Adobe Acrobat Reader is a widespread tool that is found on the desktops of many PCs and laptops (of various operating system lineage) of various industries, governmental agencies, and educational institutions. There are other viewers that are available that also read the PDF format.

4.3 NRC Requirements Met

The USNRC stated publicly at the 2008 NIRMA Conference (August 2008, Las Vegas, NV) that they do have PDF/A-based documents in the Agency Document Access and Management System (ADAMS) today. However, just complying with PDF/A does not ensure the file has all the attributes required by the USNRC. Should an electric utility/nuclear power plant seek to use PDF/A, while also complying with the USNRC e-submittal requirements, the following additional configurations would need to be applied:

- Enable “FastWeb View” (linearization)
- Compression for black and white TIFF images embedded in PDF documents is CCITT Group IV
- Documents are text-searchable (if scanned, would need to undergo an optical character recognition process)
- Image resolution is set at 300 dpi

NIRMA members are beginning to consider, implement, or convert their respective electronic record repositories to use PDFs created using the ISO 19005 PDF/A standard. In order to assist licensees using this technology, NIRMA suggests that the USNRC might consider addressing the option of using PDF/A in their next revision of their Guidance for Electronic Submissions to the NRC by specifically providing for use of the ISO 19005 PDF/A settings, with the additional settings stated above as a means to assure the NRC that electronic documents/records submitted to the NRC would comply with the E-Rule Guidance.

4.4 Implementation Guidance

NIRMA members are aware of the caution (and disclaimer cited in the ISO standard) that PDF/A only addresses a long-term **format type**. PDF/A alone does NOT guarantee long term preservation, as it is only ONE component of an entire archival strategy. NIRMA stresses that a comprehensive electronic records program not only addresses the long-term format issue, but also should include all aspects of an electronic records management program as identified in NIRMA TG15, "Management of Electronic Records."

A nuclear facility implementing PDF/A should develop a strategy and implementing procedures and processes to ensure consistent generation of PDF/A files. Additional means to ensure consistent implementation are:

- use of an automated validator
- installation of application software on a server
- training/written guidance on purpose and use of PDF/A
- contractual requirements with outside vendors for document submittals

5.0 References

- 36 CFR 1234, "Electronic Records Management"
- Association for Information and Image Management, "Frequently Asked Questions ISO 19005-1:2005 PDF/A-1", www.aiim.org/documents/standards/19005-1_FAQ.PDF
- ISO 19005-1, "Document Management – Electronic Document File Format for Long-term Preservation – Part 1: Use of PDF 1.4 (PDF/A-1)," which establishes PDF/A ("PDF-Archive")
- NIRMA TG15-2009, "Management of Electronic Records"
- PDF/A Competence Center: www.pdfa.org.
- U.S. NRC [Electronic Maintenance and Submission of Information](http://www.nrc.gov/site-help/e-submittals/10cfr1.pdf) ("E-Rule"): <http://www.nrc.gov/site-help/e-submittals/10cfr1.pdf>. Forms used to submit information electronically are available on the NRC's public Web site at <http://www.nrc.gov/site-help/e-submittals.html>.